U. S. FOREST SERVICE

RESEARCH NOTE LS-68

LAKE STATES FOREST EXPERIMENT STATION • U. S. DEPARTMENT OF AGRICULTURE

Outdoor Recreation Surveys: Length-of-Stay Bias and Its Correction by Computer

Simple random samples of recreationists taken at recreation sites may not provide representative, unbiased estimates of visitor characteristics. This problem has been pointed out before, and a computational procedure to remove the effect of the bias has been suggested.[1] A computer program to make the adjustment is reviewed in this Note.

The bias arises in on-site surveys[2] because the probability that a camper, swimmer, or other recreationist will fall in the sample and be interviewed is a function of how long he is in the campground, near the beach, and so on. Campers who stay only 2 days, for example, are only one-tenth as likely to be sampled as campers who stay 20 days. Where activities or attitudes are the main interest, this sampling of *visitor-days* may be appropriate. If the survey is made to estimate characteristics of *visitors*, however, the summarized responses must be corrected for sampling bias.

Past recreation surveys carried out on-site have not been corrected for length-of-stay bias. This is unfortunate because uncorrected surveys overestimate length of stay and all factors associated with it. For example, we know that people who live farther from a site generally visit it less often but for longer periods than do local people. Thus an on-site sample of visitors overrepresents these more distant people, who may also be better paid, more educated, have different attitudes, spend more, and so on.

No completely satisfactory way of avoiding this bias in the original data collection is available.[3] Because the bias is a simple function of length-of-stay, however, correction for it is quite straightforward. If each interview is weighted by the inverse of the length of stay, the bias disappears.[4] This weighting

---

1 *Lucas, Robert C. Bias in estimating recreationists' length of stay from sample interviews. Jour. Forestry 61: 912-914. 1963. The problem was also recognized in Standards for Traveler Studies, prepared by the Western Council for Travel Research, University of Utah, Salt Lake City, 1963, p. 23; but it was assumed that bias could only be corrected by arbitrary adjustments.*

2 *Redefining the population of interest by taking a sample from a registration list or at an area entrance or exit point avoids this bias since each recreationist can then enter the sample only once on each visit. If the target population is defined as visitors, other procedures are necessary, such as removing registrations on second and subsequent visits.*

3 *One possible way would involve first asking a respondent his intended length of stay, and then accepting or rejecting him for interviewing by some randomizing process proportionate to the inverse of his stay. Every tenth 10-day visitor could be interviewed, every fifth 5-day visitor, and so on. This would appear to work but it seems clumsy. Another would require taking a 100-percent sample (a census) of a part of the study area for a portion of the study period. This cluster sampling would be less sensitive to possible variation over time and area, however.*

4 *To be precise, the amount of time the recreationist is present during the period of the day on which sampling occurs should be used. A camper who arrives after dark and breaks camp early the next morning, for an extreme example, has zero probability of being interviewed in most surveys. However, this refinement is difficult to make in most situations, and appears to add little accuracy.*

MAINTAINED AT ST. PAUL 1, MINNESOTA, IN COOPERATION WITH THE UNIVERSITY OF MINNESOTA

scales down the overrepresented longer stays relative to the underrepresented shorter stays. Sampling with replacement is necessary; if a party falls in the sample a second time, it should be included — either by duplicating the previous interview or by reinterviewing. If it is desired to have the total number of observations after weighting be the same as the number in the original, unweighted data, a constant multiplier can be incorporated into the weighting factor.

While computing an average length-of-stay requires only a simple count of the observations for each possible stay, computing other sample characteristics that must be corrected for a length-of-stay bias involves tabulating the frequency of each response within each length-of-stay. As the number of possible responses to a question, the number of questions, and the number of interviews all increase, the mechanical procedures of listing and calculating by hand become overwhelming.

A computer program prepared at the Lake States Forest Experiment Station, designated LSWTFREQ, can more efficiently provide the type of information suggested above: weighted frequency distributions with weights determined by some critical variable, here the length-of-stay. In addition to specifying these weights, the program user can select another variable so that separate frequency distributions will be printed within each value of this "classification" variable. For example, the user might obtain a frequency distribution, corrected for a length-of-stay bias, of distances traveled for each of several recreationist income classes.

The program is general in that it can be applied to other sorts of weightings and other types of studies, provided the bias or required adjustment can be keyed to the value of some recorded characteristic of the sampled population. For example, this characteristic might be size of firm, acres of land owned, or distance from a market. In each case, the variable could be used to adjust for varying sampling intensities within the several strata defined by the values of this weighting variable.

The only critical programmed restriction is that only integer codes with a maximum value of 100 can be processed.[5] The program is written in FORTRAN 60, specifically for the CDC 1604 computer of the Numerical Analysis Center of the University of Minnesota. It can be used with minor modifications on most computers with the following characteristics:

(1) A FORTRAN 60 or FORTRAN II or compatible compiler capable of utilizing —
   (a) Two-dimensional arrays,
   (b) A variable data format (not in source program),
   (c) An input variable as the limit of an implied DO LOOP for reading the remainder of that input record;
(2) Core storage of 32,000 data words in addition to the compiler storage requirement.

A detailed description and listing of the program is available from the Biometrical Section of this Station.

---

[5] *Measurements coded by alphanumeric or other multiple punch combinations must be changed to this form.*

October 1965

ROBERT C. LUCAS
Principal Geographer
DENNIS L. SCHWEITZER
Associate Statistician